DOCUMENT RESUME

ED 387 938 EA 027 138

AUTHOR Webb, Melvin W., II
TITLE Policy Considerations in Developing Standards and Assessments for Large, Diverse School Districts.
PUB DATE Apr 95
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; Evaluation Problems; *Evaluation Utilization; Grading; Minimum Competency Testing; *Performance; School District Size; State Standards; *Student Evaluation

ABSTRACT

Although there is disagreement among educators about performance standards and their relationship to student performance, performance standards are central to the reforms under way in many states and large school districts, including Philadelphia. They are also mentioned explicitly in the Goals 2000: Educate America Act and in the re-authorization of Title I. Without a clear understanding of the issues surrounding performance standards and assessments, including their potential uses and their impact on a variety of populations, the implementation of a performance-standards and assessments-based system could harm the group that is most affected--students. This paper identifies and briefly discusses 16 policy issues that must be considered when developing a performance-based standards and assessment system. These issues include the following: (1) purposes(s) of standards/assessments; (2) method(s) of standard-setting to use; (3) types of judges to use for setting standards; (4) numbers of levels to set; (5) who determines final standards; (6) conflict between local, state, and national standards; (7) impact of standards/assessments on what is taught; (8) impact of standards/assessments on how teachers teach; (9) changes in school grading systems; (10) relationship between standards and assessments; (11) opportunity to learn; (12) impact on "special" populations; (13) school-based management versus central control; (14) "world-class" standards versus minimal competency; (15) information dissemination to the public and teachers about standards/assessment; and (16) sequence of development. (Contains 15 references.) (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 387 938

Policy Considerations in Developing Standards and Assessments for Large, Diverse School Districts

by

Melvin W. Webb II, Ed.D.
Philadelphia School District

presented to

the Annual Meeting of
The National Council on Measurement in Education

San Francisco, CA

April 20, 1995

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M.W. Webb - II

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

EA 027 138

BEST COPY AVAILABLE

INTRODUCTION

According to Baker and Linn (1993), there is little or no agreement in the psychometric community, and among educators in general, on "what performance standards are, how they are best set, and what their relationship is to details used in scoring student performance" (p. 1). Yet performance standards, and attendant performance assessments, are central to the reform efforts underway in many states and large school districts (including Philadelphia), and are explicit in the Goals 2000: Educate America Act and in the re-authorization of Title I (Linn, 1994). Without a clear understanding of the issues surrounding performance standards and assessments, including their potential uses and their impact on a variety of populations, the implementation of a performance standards and assessments-based system of curriculum reform, instructional change, and accountability could result in serious harm to those most affected by standards and assessments--students.

This paper identifies and discusses briefly sixteen policy issues that must be considered when developing a performance-based standards and assessment system. Doubtless there are others not identified here. The author encourages readers to identify additional issues that should be included.

POLICY ISSUES TO CONSIDER

The policy issues that need to be considered when developing a performance standards and assessment system include:

1. purpose(s) of standards/assessments

2. method(s) of standard-setting to use
3. types of judges to use for setting standards
4. number of levels to set
5. who determines final standards
6. conflict between local, state and national standards
7. impact of standards/assessments on what is taught
8. impact of standards/assessments on how teachers teach
9. changes in school grading systems
10. relationship between standards and assessments
11. opportunity to learn
12. impact on "special" populations
13. school-based management vs. central control
14. "world-class" standards vs. minimal competency
15. informing public/teachers about standards/assessments
16. sequence of development

This list is in no particular order, except for item 1., purpose(s) of standards/assessments, which is the first issue that any group considering their development and use should address.

DISCUSSION OF POLICY ISSUES

Purpose(s) of standards/assessments. The manner and extent to which performance standards and assessments will impact districts, schools, and students will "depend heavily on the uses to which they are put" (Linn, 1994, p.1). Linn identified four potential purposes (or uses) of performance standards and their attendant assessments; exhortation, exemplification of goals, accountability for educators, and student certification. These different uses involve different levels of "stakes" and carry different levels of risk for districts, schools, and students.

By exhortation, Linn meant the use of standards and assessments for symbolic purposes, with low stakes for individuals attached to success or failure. As an example, National Education Goal Five exhorts us, as a nation, to be "first in the world in mathematics and science achievement" by the year 2000 (NEGP, 1994).

Reaching or not reaching this standard will have limited impact on individuals.

Using standards to exemplify goals also involves low stakes for individuals. Used for this purpose, standards might provide "clear specifications of the achievement levels students are expected to attain" (Linn, 1994, p.3). The Achievement Levels developed for the National Assessment of Educational Progress are examples of standards that exemplify goals.

The use of standards and assessments as an accountability device for educators involves low stakes for students, but potentially high stakes for teachers and administrators. In Philadelphia, for example, schools that consistently under-perform relative to our standards will be "taken over" by the District, with all administrators and most teachers replaced.

The most high stakes use of standards and assessments is for student certification, which can include graduation/promotion, endorsed diplomas, special certificates, and even employment or college admissions. As an example, students who do not meet the performance standard on the New Jersey High School Proficiency Test (HSPT11) for eleventh grade do not receive a diploma, a very high stakes use of standards and assessments.

The purpose(s) for which standards and assessments will be used should be determined <u>before</u> the development process begins, because the nature of their use should guide the developmental process in terms of specificity of the performance standards and the technical rigor of the assessments.

Method(s) of standard-setting to use. Methods for setting standards on assessments vary from the somewhat simple (e.g., deciding what score-point on a scoring rubric represents proficiency) to the complex (e.g., the two-stage judgmental policy capturing method (Jaeger, 1994)). Methods also vary according to item type, with fairly well-accepted methods existing for multiple-choice items (e.g., modified Angoff and Nedelsky), and less well-accepted methods still being researched for constructed-response items (Webb and Miller, 1995). The choice of a method to use will impact the final standard that is set, as research indicates that different methods yield different results (NAE, 1993). What seems clear is that no matter which method of standard-setting is chosen, controversy will ensue. In general, the courts have upheld the use of performance standards for purposes such as certification and licensure of physicians as long as the method(s) employed to set the standards have been well-researched, well-documented, and are technically sound.

Types of judges to use for setting standards. According to Jaeger (1991), "[r]easonable results [from standard-setting] can be expected only if the judges called upon to use these methods [e.g., the modified Angoff] are highly knowledgeable of the domain in which decisions are required" (p.4), that is, experts. Deciding exactly who is an expert, however, is not an easy task. For example, is a mathematics expert a classroom teacher, a mathematics curriculum specialist, a mathematics researcher, or even a person who uses mathematics extensively in his/her work? If all these

types of individuals are mathematics experts, which one(s) should you use to set standards, how many of each type, and how should they be selected? These decisions should be related to the purpose(s) for your standards, they should be made as part of the basic design of the standard-setting process, and they might differ due to the particular political climate in your state or district. Because different types of judges are likely to produce different standards (Jaeger, 1991; NAE, 1993), the choice of judges to empanel is also bound to produce controversy.

Number of levels to set (and what to call them). As Baker and Linn (1993) asserted, one of the critical issues facing standard setters is the number of levels of performance to set and what to call them. There seems to be general agreement that more than two performance levels (i.e., more than one cut-score) are desirable, but little agreement about the optimum number. The NAEP, for example, reports against three levels (Basic, Proficient, and Advanced), while Kentucky reports against four levels, which they call Novice, Apprentice, Proficient, and Distinguished. The use of multiple levels of performance standards has the advantage of allowing students of varying skill and ability to demonstrate progress toward some standard, even if it is not the optimal standard (e.g., Proficient).

In choosing names for levels of performance, great care should be taken to avoid value-laden terms. For the NAEP, the choice of the term "Basic" for the lowest level generated controversy from the start, and continues to offend some people. The simple solution

might be to number the levels (avoid letters like A, B, C for obvious reasons), but names that have meaning attached to them like "Proficient" do have the advantage of conveying a "sound-bite" type message.

Who determines the final standards (and should they be adjusted)? The standards developed by judges are usually recommendations forwarded to some official group or person for consideration and eventual adoption or revision. For a large school district, the group making the final determination of whether to adopt the standards as is or to adjust them should include key administrators from areas such as assessment, curriculum, desegregation, language minority education, and so forth. In most districts, the Board of Education will reserve the right for final approval of the standards.

Geisinger (1991) provides guidance about factors, to consider when deciding whether or not to adjust the recommended standards. These factors include: 1) acceptable passing and failing rates, 2) the relative "costs" of classification errors [e.g., what harm will ensue from "passing" students that really aren't proficient versus "failing" students who are?], 3) organizational or societal needs, 4) adverse or disparate impact data [see "Impact on special populations" discussion below], 5) errors of measurement, and 6) errors of rating. As an example, in considering whether to adjust the final Achievement Levels for the 1992 NAEP in Mathematics, the National Assessment Governing Board considered 1), 5), and 6)

above, and decided to adjust the cut-points downward, to the lower bound of the standard error of measurement (ACT, 1993).

Conflict between local, state, and national standards. At present, we have national performance standards (e.g., The National Education Goals and the NAEP Achievement Levels), state standards (e.g., the Pennsylvania Learning Outcomes), and in many areas, local standards (e.g., the Philadelphia Standards 2000). It is inevitable that conflicts will exist between and among these standards, with one set viewed as more rigorous than the others, or as more relevant, or as more fair, and so on. In Philadelphia, we have made the policy decision that our local standards will take precedence over state and national standards. We have, however, used the national and state content standards as "templates" for the development of our local content standards, and will doubtless do the same for our performance standards. We are also looking at ways to use the NAEP and/or NAEP equatable assessments as an integral part of our total assessment package. As Mirel & Angus (1994) said "clearly articulated national content and performance standards and well-designed national methods of assessment can enhance opportunity" (p.6) for school and student improvement. We believe this is true.

Impact of standards/assessments on **what** is taught. Phyllis Aldrich recently posed an interesting, and serious, question when she asked (1994), "are we doomed to be seen as policy makers who fiddle with official curricula?". Many teachers, parents, and others view performance-based standards and assessments as exactly

that--just "us administrators" fiddling with the curriculum once again. There are legitimate concerns, however, and ones that must be considered when developing standards and assessments. Perhaps the most serious is the concern for subject areas that have traditionally been viewed as peripheral, such as fine arts, foreign language, and music, and for which we may or may not develop standards and assessments. If we don't, will these subject areas receive even less attention? If we do, will that elevate them to the same status as mathematics, science, and English? And what about new "subject" areas such as multi-cultural education? The inclusion, or non-inclusion, of controversial areas in the standards and assessment process may lead to the demise of the entire standards enterprise in a local area.

On the other hand, if the local curriculum can be positively affected by the shift to a performance-based standards and assessment system, then real change in student achievement can occur. According to Aldrich (1994), the emphasis on high standards for all students could lead teachers and administrators to conduct a realistic re-appraisal of what <u>should</u> be taught.

<u>Impact of standards/assessments on **how** teachers teach (and students learn).</u> If performance standards and assessments are going to positively impact student achievement, an intensive and ongoing commitment to staff development for teachers and time for them to experiment with new assessment methods is imperative (Resnick, 1994). Dropping a set of content and performance standards, with a new assessment system, into teachers' laps and saying "just do it"

will cause harm to teachers and students, alike. In general, current models of staff development appear inadequate for the level of training and skills upgrading necessary to enable teachers to take full advantage of these new developments. As one 9th grade mathematics teacher recently told me, "I'm not sure that I have the content knowledge and process skills called for by the NCTM Standards. How in the world am I going to teach them to my kids?"

The amount of money needed for this staff development is staggering. In Philadelphia, we recently computed the staff development costs of introducing a new standards-based assessment system into one-third of our schools, in only two grades, and only to give these teachers rudimentary training in how these new standards and assessments differed from our old curriculum and assessments plus some training in scoring constructed-response items, and concluded that the price tag was currently beyond our reach.

Changes in school grading systems. One of the benefits of a performance standards-based system is the ability to report to parents and the public on what students can do in terms of content knowledge and skills instead of merely in terms of percentiles, class rank or grade average. However, as Aldrich warns, it will be difficult for parents and the public to accept new methods of reporting student performance since "so much of what people expect of school is based on their own experience when they were in school" (p.8). As educators, we have an obligation to educate parents and the public about new grading systems and ways of

reporting test scores. If we ignore this obligation, we run the risk of alienating parents and the public. One way to address this problem might be to institute a dual reporting system (e.g., normative data and progress towards standards on the same score report) that could be slowly phased out as people grow accustomed to the new reporting methods.

The relationship between standards and assessments. When developing content and performance standards the assessment component of the system must constantly be kept in mind. Standards that cannot be effectively assessed will not help us gauge student progress. What's more, standards that may be acceptable on their face-value may be rendered unfair for students, or particular groups of students, by the way they are assessed (Phillips, 1994). In addition, assessment tasks that cannot be directly related to content and performance standards may be measuring content and skills to which students have not been exposed, a clear violation of good practice.

Opportunity to learn. There may be no thornier issue in public education today than opportunity to learn. In relation to performance standards and assessments, Debra P. vs. Turlington, a landmark testing case, provides some guidance on opportunity to learn issues. According to Phillips (1994), Debra P. requires the following:

1. that a two-to-four year notification period in advance of the implementation of a diploma test be included;

2. that information about content, format and scoring be provided to school personnel and students;

BEST COPY AVAILABLE

3. that this information provide a clear indication of the specific skills and knowledge for which students will be held accountable; and,

4. that guidelines on what constitutes acceptable performance be provided to students.

Clearly, a well thought-out system of content and performance standards can, and should, address these requirements.

Impact on "special" populations. One of the basic tenets of the standards movement is that standards and assessments are for all students (Linn, 1994). In Philadelphia, for example, our Children Achieving blueprint for systemic reform includes standards and assessments as two of the ten basic building blocks for all students achieving at high levels. As Lam and Gordon (1993) point out, though, providing equitable educational opportunities to language minority students with different cultural and linguistic backgrounds will be a major challenge. Standards and assessments may inadvertently discriminate against these students unless we recognize that linguistic barriers may inhibit their understanding of 1) the content standards, 2) the performances expected of them, and 3) the assessments with which they will be measured. In the case of assessments, we must be sure we are assessing their content knowledge and skills, and not their language ability.

In addition, initial data indicates the presence of increased adverse impact of new standards and performance assessments on historically disadvantaged students (Phillips, 1994); students who comprise the majority of students in many of our largest school districts. Portfolios, for example, may provide distinct advantages to students from non-disadvantaged families, where parental

support, community support, and educational resources are more abundant. We must be careful in the development of standards and assessments not to assume that all children have the basic necessities of life readily available to them, and penalize further those who don't.

Finally, if disabled students are to be included in the discussion, the Americans with Disabilities Act requires that we provide whatever reasonable accommodations they require for success. According to Phillips (19940, this means "that disabled students must be considered when writing goals or standards which apply to all students, when developing assessment items or tasks, and when determining passing or other reporting standards" (p.8).

Clearly, having standards and assessments apply to all students is a desirable goal; ensuring that they apply to all students in an equitable manner will be a major challenge.

School-based management vs. central control. According to Wesley Smith (1994), "lawmakers have seized standards-based reform as the tool with which to make a decentralized public school system respond to centralized policy decisions" (no page numbers in document). For local districts who have instituted school-based management, the institution of a performance standards and assessment based reform effort might be seen as a contradiction, but it doesn't have to be one. In Philadelphia, for example, we will have district-wide content and performance standards, and a district-level assessment tied to those standards, but will leave the decisions on how to move students towards our standards, that

is decisions about curriculum and instruction, up to the schools. The key will be in the level of specificity of the standards and the degree to which they are viewed by schools as being prescriptive rather than exemplary.

"World-class" standards vs. minimal competency. We currently have a number of states and districts in this country that have minimal competency-type standards in place (e.g., New Jersey) and more that are considering them (e.g., Minnesota). The hue and cry nationally and in other states and districts, however, is for "world-class" standards (whatever they are). One must assume that, in most instances, there is a considerable difference in student performance between "world-class" and minimal. Which level of performance is more equitable for students? Which level of performance will drive educational reform? If "world-class" standards are implemented, and students are held accountable to them for graduation and promotion, do they then become de facto minimal competency standards? Is there a level between "world class" and minimal that might provide a more reasonable expectation for students? These questions have not been addressed sufficiently.

Informing public/teachers about standards/assessments. As evidenced by the attacks on "outcomes-based education" in Pennsylvania and other states, many parents, teachers, and members of the general public don't really have a clear idea of what standards-based reform really is. Attention must be paid to the process of publicizing and explaining a new system of standards and assessments well in advance of its implementation. In fact, such

planning should be an integral part of the overall design of the standards and assessment development process. In providing information to various publics, as much attention should probably be paid to what the standards and assessments are not, as to what they are. You can rest assured that groups opposing standards-based reform will be spreading their message.

Sequence of development. Which should come first, the content standards, the performance standards, or the assessments? Because one of the essential attributes of standards-based reform is the seamless (hopefully) integration of standards, instruction, and assessment (Smith, 1994), the answer to this question becomes critical.

Some would argue that assessment should drive instruction, and that the development of the assessment process should come first, followed by development of the standards. The danger with this, according to Smith (1994) is that developing the assessments before the standards will lead to the assessments becoming de facto standards. Since assessments typically cannot measure everything in the curriculum we recognize as important learning, the result would be a narrowing of the curriculum--exactly what some critics of standards-based reform claim.

CONCLUSION: EXPECTATIONS FOR STANDARDS

As the preceding discussion makes clear, we have a number of expectations for standards and assessments in their role as the primary catalyst for educational reform. We expect them to lead to

fundamental change in the way teachers teach and students learn. We expect them to lead the United States into the next century with the most highly educated secondary school graduates in the world. In short, we expect standards and assessments to revolutionize our schools. However, we need to be very careful in developing our standards and assessments, in making decisions based upon their use, and in introducing them to parents, teachers, students and other publics. Otherwise, as the new century unfolds, people will look back on this period of educational reform as just another attempt to do band-aid surgery on a terminally ill patient.

REFERENCES

Aldrich, P.W., (1994) October. The impact of standards on how teachers behave in the classroom: Promises and perils. Paper at the Joint Conference for Standard Setting for Large Scale Assessments, Washington, DC.

American College Testing, (1993). Setting achievement levels on the 1992 National Assessment of Educational Progress in Mathematics: A final report. Iowa City, IA, Author.

Baker, E.L. and Linn, R.L., (1993) Winter. Towards an understanding of performance standards. CRESST Line, p. 1-2.

Geisinger, K.F., (1991) Summer. Using standard setting data to establish cutoff scores. Educational Measurement: Issues and Practice, 10(2), p.17-22.

Jaeger, R.M., (1991) Summer. Selection of judges for standard-setting. Educational Measurement: Issues and Practice, 10(2), p. 3-10.

Jaeger, R.M., (1994) April. Setting performance standards through two-stage judgmental policy capturing. Paper at the Annual Meeting of the American Educational research Association, New Orleans, LA.

Lam, T.C. and Gordon, W.I. (1993). State policies for standardized achievement testing of limited English proficient students. Educational Measurement: Issues and Practice, 11(4), p.18-20.

Linn, R.L., (1994) October. The likely impact of performance standards as a function of uses: From rhetoric to sanctions. Paper at the Joint Conference for Standard Setting for Large Scale Assessments, Washington, DC.

Mirel, J. and Angus, D., (1994) Summer. High standards for all? The struggle for equality in the American high school curriculum, 1890-1990. American Educator, p. 4-42.

National Academy of Education, (1993) July. Setting performance standards for student achievement. A report of the NAE panel on the evaluation of the 1992 NAEP Achievement Levels. Boulder, CO: Author.

National Education Goals Panel, (1994). The National Education Goals Report: Building a nation of learners 1994. Washington, DC: Author.

Phillips, S.E. (1994) October. Legal defensibility of standards: Issues and policy perspectives. Paper at the Joint Conference for Standard Setting for Large Scale Assessments, Washington, DC.

Resnick, L.B. (1994) Summer. The federal reform of education: Boon or bane for American public schools? The Journal for the Education of the Gifted, 17(4), p. 401-420.

Smith, H.W. (1994) October. Issues in standard setting: Making a decentralized system responsive to centralized policy-making. Paper at the Joint Conference for Standard Setting for Large Scale Assessments, Washington, DC.

Webb, M.W. and Miller, E.R., (1995) April. a comparison of the paper selection method and the contrasting groups method for setting standards on constructed-response items. Paper at the Annual Meeting of the National Council on Measurement in Education, San Francisco.